PROCEEDINGS *of the* FIFTH BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY

Held at the Statistical Laboratory
University of California
June 21–July 18, 1965
and
December 27, 1965–January 7, 1966

with the support of
University of California
National Science Foundation
National Institutes of Health
Air Force Office of Scientific Research
Army Research Office
Office of Naval Research

VOLUME IV

BIOLOGY AND PROBLEMS OF HEALTH

EDITED BY LUCIEN M. LE CAM
AND JERZY NEYMAN

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY AND LOS ANGELES
1967

RECOGNITION OF PATTERNS IN PERIODIC BINARY SEQUENCES

JULIAN FELDMAN
UNIVERSITY OF CALIFORNIA, IRVINE

1. Introduction

When humans are presented with information, they tend to search for structure in the information—to encode it, to organize it, to look for regularities. The existence of this structure-seeking behavior can be seen in the behavior of subjects predicting a random sequence of binary events and in the behavior of subjects predicting more highly structured sequences. (A *random sequence* is a colloquialism for a sequence generated by a mechanism which for two events, 1 and 0, has $P\{1\} = \pi$ and $P\{0\} = 1 - \pi$. A more *highly structured sequence* contains additional constraints on conditional probabilities. *A completely deterministic sequence* is generated by rules of the form "$S_i \to 0, 1$" where the S_i form a mutually exclusive and exhaustive set of states for the events in the sequence, that is, each event follows one and only one state.) Although it is quite clear that subjects search for structure, the details of this structure-seeking behavior continue to elude investigators. In pursuit of models of structure-seeking behavior, investigators have studied behavior on completely deterministic sequences as well as on random and more highly structured sequences. The focus of the present paper will be on models of behavior in experiments using completely deterministic sequences. Two models proposed by other investigators will be reviewed and a third model will be proposed.

2. The stimulus-pattern model

In the typical experimental situation of concern here, the subject is asked to predict each symbol in a sequence of binary events. The event sequence consists of repetitions of a basic period, for example, $101010 \cdots$, $110010110010 \cdots$. After each prediction, the subject is informed of the event. Thus the experiment consists of an alternation of predictions and events—$P_1E_1P_2E_2P_3E_3 \cdots$. The subject continues to make predictions until he reaches a criterion, for example, a number of consecutive correct predictions equal to twice the length of a period.

In the stimulus-pattern (SP) model (Kochen and Galanter [4]) the subject is depicted as learning the sequence by learning the conditional relations that define the sequence. For example, to learn the sequence $1010 \cdots$, the subject learns that $1 \to 0$ and $0 \to 1$. More complex sequences are learned by learning

higher-order conditional relations, after the lower-order ones fail, for example, $110110 \cdots$ requires learning the second-order conditional relations $11 \rightarrow 0$, $10 \rightarrow 1$, $01 \rightarrow 1$.

After developing this model, Kochen and Galanter ([4], pp. 284–285) ". . . tried the experiment with one highly sophisticated S[ubject]. The S was, moreover, permitted to keep a written record of acquired information and knew that he was searching for a periodic pattern. The S was tested on . . . seven sequences, each of which was repeated ten times. The order in which the 70 sequences were selected was randomized. The smallest number of trials at which the correct pattern was first suspected is indicated in table [I] for 10 repetitions of each of the sequences. (The predictions of the SP model are also in table I.)

TABLE I

OBSERVED NUMBER OF TRIALS TO "DISCOVER" THE PATTERN IN PERIODIC SEQUENCES
(Source: table III, [4])
Note: the number marked by asterisk should be 3 (JF).

Sequence	Median Behavior	Prediction of SP Model
01	4	4*
001	5.5	5
0011	5.5	6
01001	7.5	8
000111	7	8
111010	8.5	9
1100111100	15	15

This pilot experiment, hardly to be considered conclusive evidence for the proposed program [model], does nevertheless lend a certain plausibility to such a program [model]."

Kochen and Galanter also tried to fit the stimulus-pattern model to data previously reported by Galanter and Smith [2]. In the Galanter and Smith experi-

TABLE II

TRIALS TO SOLUTION IN GALANTER AND SMITH EXPERIMENT
(Source: table I, [2] and table IV, [4])
Note: the asterisks indicate that predictions of the model were not supplied.

Sequence	Median Behavior	Prediction of SP Model	Prediction of Modified SP Model
01	5.5	4	7
001	14.0	5	14
0011	14.0	6	14
0001	14.0	*	*
00001	26.5	*	*
01001	49.0	8	50

ment, a separate group of 15 to 20 subjects predicted each of six sequences. The subjects were instructed along the following lines: "I am thinking of a *zero* or a *one*. I want you to guess which it is. After you have guessed, I will tell you what it is, and then you are to guess again. You are to try to anticipate what I will say each time." The description of the experiment is again unclear on the details of the method used to present the sequence and on the criterion of solution. The median number of trials to solution is presented in column 2 of table II. The predictions of the basic SP model are given in column 3 of table II. Kochen and Galanter suggested that since the subjects of the Galanter and Smith experiment were not permitted to keep a written record, the model should be modified to take this procedural difference into account. Kochen and Galanter presented revised predictions for four sequences (column 4 of table II). Although the fit appears somewhat better, the modifications required to generate these fits are not uniform, that is, a different parameter is required for each of the four sequences.

3. The run-structure model

In this model, the subject is depicted as recoding the binary sequence into a sequence of integers representing the run lengths (for example, $1010 \cdots \rightarrow 1:1$ and $11001100 \cdots \rightarrow 2:2$), and learning the recoded sequence as a serial list (Keller [3]). The run-structure (RS) model derives from the work of Restle and others who have emphasized the importance of runs of events as stimuli in binary-choice experiments with random and structured sequences. The strong effect of the number of runs in the period on the number of trials required to learn completely deterministic sequences can be seen in the Galanter and Smith data (table II). Hence evidence for the plausibility of the RS model is available from behavior on random, structured, and completely deterministic sequences.

To obtain more conclusive evidence of the importance of the number of runs in the period (or the code length of the period) on the behavior of subjects predicting completely deterministic sequences, Keller conducted a lengthy study in which he manipulated period length, code length, and the number of different runs in the period. The number of different runs in the period is the number of runs, that is, the code length, less any duplicates (any runs of the same symbol of the same length) for example, $111001100100 \cdots$ has period length twelve, code length six, and four different runs. The periods of Keller's sequences were generated by combining period lengths of six, nine and twelve with code lengths of two, four, and six. The 32 periods used are indicated in table III. Each subject predicted only one sequence, and each sequence was predicted by four subjects. The subjects were asked simply to predict; they were not told that the sequence contained a pattern but only that their performance should improve as the experiment progressed. The events were presented via light bulbs on a panel in front of the subject. Only three 0's were used in the first run for sequences of code length two. Ten subjects failed to meet the criterion of no errors in a sequence of trials equal to three period lengths.

TABLE III

MEAN TRIAL OF LAST ERROR IN KELLER EXPERIMENT
(Source: tables 1 and B1, [3])

Sequence	Mean	Sequence	Mean
1. 000011	47.25	17. 010000101	172.25
2. 000011	19.25	18. 001001011	289.75
3. 000001	26.50	19. 000100101	167.50
4. 000001	60.75	20. 001100101	133.50
5. 001101	98.75	21. 000000001111	63.75
6. 001101	113.25	22. 000000000111	91.25
7. 010001	69.50	23. 000000000011	90.50
8. 010001	130.75	24. 000000000001	63.25
9. 000001111	72.75	25. 000110000011	142.25
10. 000000111	42.75	26. 010000000001	78.75
11. 000000011	64.75	27. 000100000011	238.25
12. 000000001	54.75	28. 001111000001	104.25
13. 001100011	51.25	29. 001000001001	186.00
14. 010000001	58.75	30. 010000000101	106.25
15. 000100111	129.25	31. 001100011011	310.00
16. 001000011	109.25	32. 000100001001	110.00

The effects of code length and period length in Keller's data are summarized in table IV. As period length increased, the mean trial of the last error increased

TABLE IV

MEANS FOR TRIAL OF LAST ERROR IN KELLER EXPERIMENT
(Source: table 3, [3])

Period Length	Code Length 2	4	6	Total
6	38.44	103.06		70.75
9	58.75	87.12	190.75	112.21
12	77.19	140.88	178.06	132.04
Total	58.12	110.35	184.41	109.28

regularly for code length two. In code lengths four and six, inversions were obtained. As the code length increased, the number of trials to criterion increased for each period length. While the data were regular for each period, some interesting overlapping occurred, for example, for period length six and code length four, the mean was 103.06, while for period length nine and code length two, the mean was 58.75. The cause of the inversions might have been the increased variability in behavior at longer code lengths. Keller also showed that the number of different runs had a regular effect. Subjects required more trials to learn periods containing four different runs than three different runs.

In addition to presenting the foregoing analyses of his data which are consistent with the RS model he presented, Keller analyzed his data for evidence of

agreement with the SP model. He did not find any orderly relation between the conditional structure of the sequences and the behavior of his subjects. Keller also examined a variant of the SP model which uses variable length stimuli. For example for the period 00001, the variable length stimuli and the associated responses would be $1 \rightarrow 0$, $10 \rightarrow 0$, $100 \rightarrow 0$, $1000 \rightarrow 0$, $10000 \rightarrow 1$, the comparable equal length stimuli would be $0000 \rightarrow 1$, $0001 \rightarrow 0$, $0010 \rightarrow 0$, $0100 \rightarrow 0$, $1000 \rightarrow 0$. Keller did not find any orderly relation between stimulus length as represented in this model and the difficulty of the sequence.

4. The stimulus-pattern-response-string model

A third model, the stimulus-pattern-response-string (SPRS) model is a development of the application of a SP model to a structured sequence (Feldman and Hanna, [1]). Although the results of this application were surprisingly successful, the manner in which subjects reeled off a string of responses indicated that a modification of the SP model might be appropriate.

The general form of the SPRS model resembles the SP model—both models consist of a set of conditional relations of the form $S \rightarrow R$. The principal difference between the models is in the structure of R. In the SP model, R is a single response. In the SPRS model, R is a string of responses. The development of either model for a particular sequence begins with the development of a conditional relation with S a 0 or 1 and R a 0 or 1. The failure of either model to make a correct prediction results in a modification of S and the addition of a new $S \rightarrow R$ relationship. In the SP model, each response is generated by matching the preceding events with a stimulus S; the appropriate response is the one associated with the stimulus. In the SPRS model each stimulus has associated with it a string of responses, and a response can be generated either as in the SP model or by selecting the next response on the response string. In general after an incorrect prediction, there is a search for a new stimulus; after a correct prediction, the next response is obtained from the response string. When the response string is exhausted, a new stimulus is found and the response string of that stimulus is used.

One way of contrasting the SP model and the SPRS model is that the SP model grows by modifying S-strings or by adding new S-strings, while the SPRS model can modify both S and R strings. For example, the SP model develops a representation of the sequence 11001100 $\cdots$ in the following fashion. First, it establishes that $1 \rightarrow 1$. When that fails on trial 3, the model modifies $1 \rightarrow 1$ to $11 \rightarrow 0$. Similarly $10 \rightarrow 0$, $00 \rightarrow 1$, and $01 \rightarrow 1$ are developed. On the same sequence the SPRS model will develop the string 11 and then the string 00. These two strings will then be joined into the string 1100.

A first version of the SPRS model (see appendix) has been used to predict the set of 26 periodic event sequences listed in table V. This set of sequences was developed by examining all binary sequences of lengths two, three, four, five, and six and eliminating sequences which could be obtained from other sequences

by the operations of rotation and/or complementation. For example, 0100 can be obtained from 1000 by rotation—repeated circular shifting of the sequence to the right—100 can be obtained from 011 by complementation—replacing 1's with 0's and 0's with 1's. 1011 can be obtained from 0001 by applying both operations. This procedure provided a list of twelve basic periods for the stimulus sequences: 01, 001, 0011, 0001, 00001, 00011, 00101, 000111, 000011, 000001, 000101, 001101. Each of these basic periods or its complement was used as the period of a sequence. A rotation of the basic sequence or a complement of the rotation was the period of another sequence except that 01 was only presented once. The addition of sequences with periods 00001111, 11000011, and 10111 completes the list of 26 sequences.

These sequences were presented by Feldman and Janet C. Cornsweet to 40 subjects in a 2×2 design. One dimension was the order of presentation. The first order of presentation was constrained so that if the basic sequence or its complement was in one half of the list, the rotation or its complement was in the other half of the list. If the first order is $S_1, S_2, S_3, \cdots, S_{26}$, then the second order is the reverse of the first—$S_{26}, S_{25}, S_{24}, \cdots, S_1$. Half of the subjects were given the sequences in this first order, and the other half were given the sequences in the second order. The second dimension was availability of history. Half of the subjects (the history condition) had available to them a printed record of the events preceding the event to be predicted; the other half of the subjects (the no-history condition) did not have such a record available to them.

An analysis of the behavior of these 40 subjects indicated that order and history had the expected effects. Subjects did better on the sequences presented later than the ones presented earlier. The principal deviation from this general finding was for sequences with code length two. This raises the question whether experience with more complex sequences does not hamper the subject's ability to

TABLE V

MEDIAN TRIAL OF LAST ERROR IN FELDMAN AND CORNSWEET EXPERIMENT

Sequence	No History	History	Model	Sequence	No History	History	Model
1. 01	3.0	3.0	3	14. 10101	7.5	6.0	19
2. 001	6.0	5.5	4	15. 111110	7.0	6.0	7
3. 101	9.5	5.0	4	16. 001000	14.5	8.0	8
4. 1110	5.0	5.0	5	17. 000011	7.0	7.0	7
5. 0100	9.0	5.0	5	18. 110011	8.5	8.0	8
6. 1100	4.0	5.0	5	19. 000111	6.0	5.0	7
7. 0110	7.0	5.0	5	20. 110001	9.5	8.0	8
8. 00001	6.0	6.5	6	21. 111010	10.5	6.0	12
9. 10111	9.5	6.0	6	22. 010100	11.0	9.0	19
10. 11011	10.5	8.0	7	23. 110010	12.0	8.5	19
11. 11100	5.0	6.0	6	24. 011010	21.0	8.0	32
12. 00110	9.0	8.0	7	25. 00001111	8.0	7.0	9
13. 00101	8.5	5.5	10	26. 11000011	13.5	9.0	10

recognize simpler sequences. In general, the median trial of the last error for subjects who had the preceding events available to them was less than for the subjects who did not have a history available. On the 26 sequences, the history subjects did better on 20, the no-history subjects did better on 3, and there were 3 ties. Again the contrary sequences were of code length two (see table V).

The experiment also provided some information on the effect of the two characteristics of periods considered above, period length and code length, and a third characteristic, rotation. The length of the period appears to be a lower bound on the trial of the last error. Code length had the same effect reported in the Keller and Galanter and Smith experiments. The period characteristic unique to this experiment was rotation. Of the 13 comparisons in the no-history condition, only one violates the anticipated effect that rotated sequences have a higher number of trials to criterion than basic sequences. In the history condition, there are two ties and three inversions in the 13 comparisons (see table V). This strong result on the effect of rotation is quite significant for the SP model described above. That model does not distinguish between basic sequences and rotated sequences. Subjects clearly do. Thus, the Feldman and Cornsweet data also tend to contradict the simple SP model.

The trial of the last error of the SPRS model is presented in column 4 of table V. The average discrepancy between the behavior of the model and the behavior of both the no-history and the history conditions is the same, 3.04 trials. This is not a particularly impressive statistic. However, four sequences 10101, 010100, 110010, 011010 account for a large part of this discrepancy. In the no-history condition these four sequences account for 37.5 of the 79.0 discrepant trials. The average discrepancy for the remaining 22 sequences is 1.9. In the history condition, these four sequences account for 57.5 of the 79.0 discrepant trials. The average discrepancy for the remaining 22 sequences in the history condition is 0.98 trials.

5. Discussion

While the experimental evidence on the recognition of periodic patterns in binary sequences contains large variations, the evidence is consistent in direction with intuitive expectations. The number of trials required to recognize a pattern is directly related to the length of the period, the number of runs (code length) in the period, the presence of noise in the form of part of a run at the beginning of the sequence (rotation), and the inexperience of subjects with similar sequences. Perhaps the only surprise in the data is the strength of the effects of code length and rotation.

Although the factors affecting recognition of periodic patterns in binary sequence are readily identifiable, the identification or synthesis of a satisfactory model of this behavior has been much less tractable. First, none of the models that have been suggested contains an adequate model of memory, and all of the models provide better fits to data obtained where subjects have the previous

history of the sequence available to them. Nevertheless, how well do the models fit the history data and the general characteristics of the no-history data?

The naive model underlying all of the experimental and theoretical work on this problem is the period-length model. According to this model, the subject will make his last error on trial p where p is the length of the period. While this model (or rather a slight variant which says that the last error is on trial $p + 1$) provides a close fit to the data (the average error on the Feldman and Cornsweet no-history condition is 1.95 trials for 22 sequences, and the corresponding statistic for the history condition is .59 trials), the model has certain serious flaws: it cannot predict the observed effect of code length and rotation.

Before presenting their SP model Kochen and Galanter [4] discussed the λ model, a variant of the period-length model proposed by Bush. In the λ model, the pattern used to predict event t of the sequence is the smallest period consistent with events 1 through $t - 1$. The tentative period length is λ. While the λ model was rejected by Kochen and Galanter as psychologically unrealistic, the fit of the λ model to the Feldman and Cornsweet data is quite impressive. The average error on the Feldman and Cornsweet no-history condition is 1.98 for 22 sequences, and the corresponding statistic for the history condition is .57 trials. The major difficulty with the predictions of the λ model (as opposed to the reservations about the assumptions) is the inability of the model to predict the observed effect of code length. The λ model is consistent with rotation effect.

The first of the models described above, the SP model of Kochen and Galanter, is intuitively appealing. The fit to the single subject data presented by Kochen and Galanter is impressive. While the discrepancy between the SP model and the Keller and Galanter and Smith data may be largely a function of the absence of a memory model, the qualitative differences are discouraging. The SP model is also unable to account for the rotation effect observed by Feldman and Cornsweet.

The RS model of Keller appears promising. It is unfortunate that the model is not completely specified so that it could generate trial-by-trial predictions. The Keller model resembles the Simon and Kotovsky [5] model of pattern recognition and the Simon and Kotovsky model might be adapted to investigate Keller's hypotheses.

The version of the SPRS model, presented in this paper is, like its competitors, a better predictor of history conditions than no-history conditions. While there are a few large discrepancies for certain patterns between the Feldman and Cornsweet data and the model, the SPRS model is more consistent with the qualitative features of the data—effect of code length and rotation—than any of the other models discussed. Furthermore, the SPRS model is consistent with other observations on the differences between behavior after correct and incorrect prediction and on the tendency of subjects to make long strings of predictions when permitted to do so. Thus the SPRS model appears promising.

While we do not yet have a complete understanding of the processes underlying the recognition of patterns in binary sequences, the empirical and the-

oretical efforts that have been made are encouraging and point to several additional areas of investigation. The large variances that have been reported suggest further study of individual differences and the parameters associated with individual behavior. Of particular interest are the biases or initial sets that subjects bring into the experimental situation. A second major area for further study is the development of models with the ability to treat hypotheses (about the structure of the sequence) as entities. While some of the models discussed above suggest that the important units are larger than a single symbol, all of these models are far less sophisticated information processors than the subjects they purport to represent.

The SPRS program was written in SNOBOL and executed at the Western Data Processing Center, UCLA, utilizing a terminal at Irvine and the WDCOM teleprocessing system. The WDPC staff was most generous in permitting me to utilize a preliminary version of the WDCOM system.

The author is indebted to Dr. L. Keller for permission to reprint material from [3].

The research reported here has been supported in part by a grant from the Carnegie Corporation of New York.

◇ ◇ ◇ ◇ ◇

APPENDIX

Description of SPRS *program, version* 1. The basic premise of the SPRS program is that the subject is trying to construct a model of the event sequence. The model is a string of symbols representing a hypothetical period. The ultimate model of a periodic sequence is the string of symbols that is the period.

The *event sequence* is represented by a string of E's: $E_1E_2E_3 \cdots E_{t-2}E_{t-1}$. The *model* is represented by a string of S's: $S_1S_2S_3 \cdots S_n$.

The program uses the model and the events preceding trial t to generate a prediction for trial t. The program also modifies the model to make it a more adequate representation of the period and hence a better predictor of the sequence.

The program is represented in the flowchart of figure 1. The program begins by selecting a *model* from its list of models. The model selected is the first model on the list of models for which the first symbol of the model S_1 is the same as the last event of the sequence E_{t-1}.

After the model is selected, the program partitions the model into a *stimulus string* and a *response string*. The stimulus string is that part of the model beginning with S_1 and ending with S_i that matches symbol-for-symbol a part of the event sequence beginning with E_{t-i} and ending with E_{t-1}. The response string begins at S_{i+1} and ends at S_n. This matching is achieved by matching E_{t-2} with S_1 and E_{t-1} with S_2. If these symbols match, then an attempt is made to match E_{t-3} with S_1, E_{t-2} with S_2, and E_{t-1} with S_3, and so forth. Thus if the event

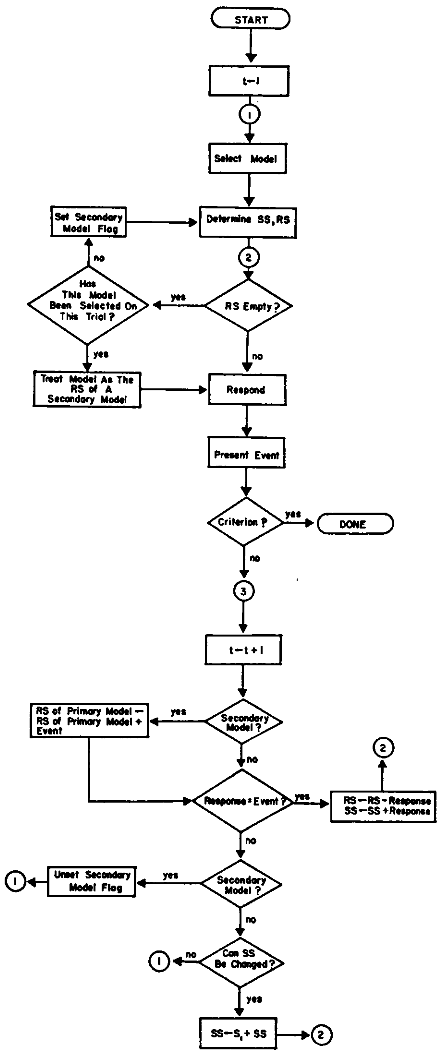

FIGURE 1

Flowchart of SPRS-1.

sequence is $\cdots$ 111000110111 and the model is 110, then the matching procedure would partition the model into the stimulus string 11 and the response string 0.

If the response string is empty and the model has just been selected, then the model is treated as the response string of a *secondary model.* If the response string is empty and the model has been selected on a previous trial, then the program selects a *secondary model.*

If the response string is not empty, the program selects the first symbol on the response string as its response for trial t.

The tth event is then presented by a subprocess representing the experimenter. The experimental procedure now checks to determine whether the program has reached the criterion: a successive number of correct responses equal to twice the length of the period of the event sequence. If the criterion has been reached, the experiment moves on to the next sequence or *terminates*. If the criterion has not been reached, the experiment continues.

The trial number is increased by one.

If the model was a secondary model, the response string of the primary model is augmented with the last event, that is, S_{n+1} becomes E_{t-1}.

If response$_{t-1}$ was equal to event$_{t-1}$, then the partition of the model is moved over one symbol to the right. The last prediction is added to the stimulus string and taken away from the response string. The program then returns to the task of generating the next response.

If the response was not equal to the event and the model was a secondary model, the program starts over again with the selection of a new primary model.

If a primary model is predicted erroneously, then the program attempts to modify the stimulus string. The stimulus string will be augmented by prefacing it with an S_0 equal to S_1 if the primary model predicted the end of a run prematurely, that is, $E_{t-1} = E_t$ and response$_t \neq E_t$. If the primary model cannot be modified, then the program starts over again with the selection of a new primary model.

Example. The event sequence is 111110111110111110 $\cdots$.

Trial 1. The program begins with a model list consisting of two models, 1 and 0. On the first trial, there are no previous events, the model does not make a response. The first event is 1.

Trial 2. The program selects the model represented by 1. However, matching the one symbol of the model to the first event results in an empty response string. The model is then treated as the response string of a secondary model and the response 1 is generated. The event is 1. The criterion has not been reached.

Trial 3. A 1 is added to the primary model. The response was correct. The program looks for a response on response string of the secondary model which is now empty. (It was a 1 on trial 2.) The program searches for a secondary model and finds the primary model. The primary model is 11 and the secondary model is 11. The secondary model is treated as a period and generates the response 1. Event 3 is 1.

Trial 4. A 1 is added to the primary model which is now 111. The criterion has not been reached. The response was correct. The program uses the second 1 on the secondary model 11 for the response for trial 4. Event 4 is 1.

Trial 5. A 1 is added to the primary model which is now 1111. The criterion has not been reached. The response was correct. The response string of the secondary model has been exhausted. A new secondary model 1111 is selected. It generates a response, 1. The fifth event is 1.

Trial 6. A 1 is added to the primary model which is now 11111. The criterion

has not been reached. The response was correct. The response is 1. The event is 0.

Trial 7. A 0 is added to the primary model which is 111110. Incorrect prediction. The present models are abandoned. A new model, 0, is selected. Its response is 0. The event is 1.

Trial 8. A 1 is added to the primary model, 01. Incorrect prediction. The new model is 111110. The stimulus string is 1. The response string is 11110, and so forth.

REFERENCES

[1] J. Feldman and J. F. Hanna, "The structure of responses to a sequence of binary events," *J. Math. Psychol.*, Vol. 3 (1966), pp. 371–387.

[2] E. H. Galanter and W. A. S. Smith, "Some experiments on a simple thought-problem," *Amer. J. Psychol.*, Vol. 71 (1958), pp. 359–366.

[3] L. Keller, "Run structure and the learning of periodic sequences," unpublished Ph.D. thesis, Indiana University, 1963.

[4] M. Kochen and E. H. Galanter, "The acquisition and utilization of information in problem solving and thinking," *Information and Control*, Vol. 1 (1958), pp. 267–288.

[5] H. A. Simon and K. Kotovsky, "Human acquisition of concepts for sequential patterns," *Psychol. Rev.*, Vol. 70 (1963), pp. 534–546.